

## Article

# Teaching Statistics in Biology: Using Inquiry-based Learning to Strengthen Understanding of Statistical Analysis in Biology Laboratory Courses

Anneke M. Metz

Department of Cell Biology and Neuroscience, Montana State University, Bozeman, MT 59717

Submitted July 12, 2007; Revised November 26, 2007; Accepted April 15, 2008  
Monitoring Editor: Diane Ebert-May

There is an increasing need for students in the biological sciences to build a strong foundation in quantitative approaches to data analyses. Although most science, engineering, and math field majors are required to take at least one statistics course, statistical analysis is poorly integrated into undergraduate biology course work, particularly at the lower-division level. Elements of statistics were incorporated into an introductory biology course, including a review of statistics concepts and opportunity for students to perform statistical analysis in a biological context. Learning gains were measured with an 11-item statistics learning survey instrument developed for the course. Students showed a statistically significant 25% ( $p < 0.005$ ) increase in statistics knowledge after completing introductory biology. Students improved their scores on the survey after completing introductory biology, even if they had previously completed an introductory statistics course (9%, improvement  $p < 0.005$ ). Students retested 1 yr after completing introductory biology showed no loss of their statistics knowledge as measured by this instrument, suggesting that the use of statistics in biology course work may aid long-term retention of statistics knowledge. No statistically significant differences in learning were detected between male and female students in the study.

## INTRODUCTION

In *Bio2010: Transforming Undergraduate Education for Future Research Biologists*, the National Research Council (NRC) recommends an increased emphasis on interdisciplinarity in biology curricula, including more emphasis on mathematics and statistics (NRC, 2003). Little work has been done on the teaching of statistics in an integrated manner in undergraduate biology courses. Horgan *et al.* (1999) describe continuing education workshops in statistics for biological research scientists, but they provide only anecdotal evidence of student learning. A'Brook and Weyers (1996) surveyed the level of statistics teaching in undergraduate biology degree programs (in the United Kingdom), and they found that statistics course work is almost invariably separated from biology course work. Both Horgan and A'Brook conclude that biology students are not learning statistics at the undergraduate (and even graduate) levels, which results in the

lack of, or misused, statistical analysis by many professional biologists.

The experience of our faculty teaching undergraduate biology mirrors the observations of both A'Brook and Weyers. In our traditional curriculum, statistics was a required course for the major, but it was not a prerequisite for any particular biology course. As a result, students took the course at greatly varying points in their curriculum, usually as juniors or seniors. Furthermore, because understanding of statistics was not consistently or uniformly stressed within the biology curriculum, students made few intellectual connections between their general statistics course (which uses examples from a wide range of fields, including economics, agriculture, medicine, sociology, and engineering) and the use of statistics in biology. Anecdotal evidence from our courses also suggested that biology students had difficulty understanding quantitative information presented graphically or in tables, were unable to manipulate raw data sets, and could not interpret statistical measures of significance such as  $p$  value. The Montana State University (MSU) biomedical sciences major attracts a high percentage of students

DOI: 10.1187/cbe.07-07-0046

Address correspondence to: Anneke M. Metz (anneke@montana.edu).

who hope to attend graduate, medical, or professional school; yet, even this highly motivated group of students seemed to have few skills in analyzing and manipulating biological data, particularly at the lower-division level, even after completing a formal statistics course.

The literature suggests that this is not an uncommon occurrence in biology programs. Peterson (2000) describes an introductory biology course where students were required to perform an inquiry-based experiment to examine the relationships between macroinvertebrates in a campus nature preserve. Instructors in that course found that students had to first be taught statistical analysis before they could analyze their own data. This instruction required time beyond what was available in the classroom, and students continued to struggle with differentiating between statistical significance of data and the significance of their findings with respect to the experimental objective. Using several pre- and postproject exam questions, Peterson reported statistically significant improvement in basic knowledge of statistics in only one of two sections that incorporated supplemental statistics instruction.

Maret and Ziemba (1997), in their investigation of teaching students to use statistics in scientific hypothesis testing, noticed that students, especially lower-division undergraduates, have few opportunities to practice the use of statistics in testing scientific hypotheses. They suggest that students should be taught the use of statistical tests early in their undergraduate careers, and concomitantly be given opportunity to practice applying statistics to test scientific hypotheses. To ensure that biology majors will have a stronger understanding of quantitative and statistical analysis, we have begun reforming our biomedical sciences curriculum to better integrate mathematics and biology, and in particular statistics and biology.

Although the teaching of statistics in the biology classroom is understudied, we can look to lessons learned in statistics pedagogy to reform biology curricula. Modern introductory statistics courses were established in the late 1970s, and they have seen a continual growth in enrollment since that time (Guidelines for the Assessment and Instructions in Statistics Education [GAISE], 2005). The growth in enrollment has resulted in a democratization of the statistic student population—students are coming into statistics with increasingly diverse backgrounds and professional goals. This has meant shifting from teaching a narrowly defined set of professionals-in-training (e.g., scientists) to making statistics relevant and accessible to students with a wide range of career interests and quantitative skills (Moore, 1997; GAISE, 2005). Statistics academicians have responded with a number of curricular reforms that are useful to consider.

Moore (1997), in considering how to best craft introductory statistics courses for modern students who will become citizens and professionals, but (mostly) not professional statisticians, suggests that “the most effective learning takes place when content (what we want students to learn), pedagogy (what we do to help them learn) and technology reinforce each other in a balanced manner.” There are several key points to this statistics reform. One, it is clear that active learning, with components including laboratory exercises, group work, work with class-generated data, and student written and oral presentations, increases student enthusiasm for, and learning of, statistics. Two, statistics

instructors indicate that students are more motivated and better understand concepts when real (and especially student-gathered) data sets are used. It is also important to use technology (e.g., statistics software) to emphasize statistical literacy rather than tedious calculations (Moore, 1997; Garfield *et al.*, 2002; GAISE, 2005).

From the literature and our own experience, it was clear that to truly integrate statistics and biology and engage students in the use of statistics in their major program, we would need to provide students with formal training in statistics as early as possible, give them the opportunity to practice the use of statistics, and demonstrate the importance of statistical data analysis in biological research. At MSU, students take a centralized introductory statistics course with peers in many different major disciplines. As part of an ongoing curricular reform effort, we shifted the biomedical science major requirements so that our students now take introductory statistics as incoming freshmen, rather than in their junior or senior year. We then designed our introductory biology course, taken by our majors immediately after completion of introductory statistics, to incorporate the best practices of statistics teaching reform.

We designed and piloted an instrument to determine whether reinforcement of statistical learning (by practice in using statistics in biological applications) aids in student retention of material learned in formal statistics course work, and increases student understanding of statistical concepts. Although assessments exist to measure learning in statistics curricula (e.g., the Statistical Reasoning Assessment, Garfield, 2003), there is no instrument available for measuring learning of statistics in a biology curriculum. The survey was administered to students at the beginning and end of introductory biology course. We also resurveyed students 1 yr after completion of introductory biology to determine whether students retained their knowledge of basic statistics in a biological context. Finally, we examined the impact of incorporating biology examples into introductory statistics, and we investigated whether there were gender differences in learning within our study cohort.

## METHODS

### *Curriculum*

We sampled two cohorts of biomedical science majors (2005 and 2006 incoming classes) for this study. Biomedical science students take introductory statistics (Statistics 216), taught by the mathematical sciences department, in the first semester of their freshman year, as a prerequisite to biology course work. In fall 2005, a special section of introductory statistics was offered for biology students that emphasized the use of biological examples in teaching the course material. The course was otherwise identical to other sections of introductory statistics offered that semester. Sixteen students who went on to enroll in the spring 2006 introductory biology course (Biology 213) completed this special section. Forty-four additional students enrolled in regular introductory statistics sections in fall 2005. The remaining students in 2005 satisfied their statistics prerequisite at other institutions, took statistics concurrently, or did not satisfy the prerequisite before enrollment in introductory biology. The special statistics section for biology majors was not available to the 2006 cohort. Over two semesters (spring 2006 and 2007), 264 (123 male, 141 female) students in total completed introductory biology.

## Instructional Methods

In introductory biology, students were given a brief review of some elements of basic statistics (charting categorical observations, mean/median/mode, analyzing difference between two populations, confidence interval, distinguishing correlation from causation,  $p$  value), and data graphing (scatterplot, box plot, and histogram). In the laboratory component of the course, students were provided with an introduction to the Minitab statistical software package (Minitab, State College, PA), and they completed a measurement, data collection, and statistical analysis exercise (the "Shell Lab," described below) at the beginning of the course. This material was reinforced with a statistics primer, an appendix in the laboratory manual that reviews basic statistical principles and provides a guide to the Minitab program. For the remainder of the semester, students completed six additional inquiry-based lab exercises, requiring them to collect data and perform basic statistical analyses such as comparison of the means of two populations, distribution of measurements in a population, comparison of actual allele frequencies with expected frequencies, and regression analysis of the association between two variables. The laboratory exercises were designed in a similar manner to the "teams and streams" model (Luckie *et al.*, 2004), in which 2-wk-long laboratory blocks combine short introductory exercises with modules where students design their own experiments.

## Shell Lab

Students, working in groups of four, performed three activities using a variety of sea shells (spotted arks, ear moon shells, tiger moon shells, clams, Caribbean arks or brown stripe shells; available in bulk from Seashells.com) and peanuts in the shell. This lab is a variation of the "Find Your Peanut" inquiry exercise from Duke University's Center for Inquiry-based Learning (Budnitz and Guentensberger, 2001), and it included three activities. Activity 1 (demonstrates measurement variation): Students were instructed to "measure the length of 20 peanut shells." All student groups measured the same set of peanut shells. Data sets for all the groups were compared, and variations in data measurements were discussed. Activity 2 (correlation of quantitative variables, use of scatterplot): Each group received 10 specimens of one type of shell, and they were tasked to determine whether "height" and "width" (self-defined by students) correlated or varied independently. Activity 3 (comparison of two populations): Each group received two sets of 20 shells, from "warm water" and "cold water," and they had to determine whether there was a statistically significant difference between the two sets. Students were free to choose their measurement strategy to determine "difference." Data collection and anal-

ysis were performed using Minitab statistical software. This exercise trained students in some basic statistical analyses techniques that they then used in the remaining labs for the semester.

## Statistics Survey Design

To determine the level of incoming statistical knowledge for students entering introductory biology, we developed a statistics survey (available as Supplemental Material) consisting of 11 multiple-choice and short-answer items (Table 1). The instrument included items at four (knowledge, application, analysis, and evaluation) of the six levels of Bloom's taxonomy of learning objectives (Bloom, 1956). Items were chosen based on statistical concepts both taught in introductory statistics and used in data analysis in the introductory biology labs. We used a multiple-choice (with one short answer) format to allow the survey to be completed as a student activity in the lecture component of the course. Preliminary instrument validation was performed by three introductory biology instructors at MSU and two outside content experts. These initial reviews of the statistics instrument provided useful feedback on question wording and accuracy of answer choices.

## Delivery of the Survey

The survey was administered on the first day of class to students in introductory biology ("presurvey"). Presurveys were not returned. We asked students to voluntarily self-report class rank, gender, and previous/concurrent statistics course work on the survey. To gauge student learning of statistics by the end of the course, the identical instrument was readministered to the students in a class period during the last week of the course ("postsurvey"). Students were not told ahead of time that the survey would be readministered.

One hundred eighty-six of 264 (70.5%) introductory biology students completed both the pre- and postsurveys and made up the study sample. Of the 186 students, 87 were male and 99 were female. Students were categorized as having completed statistics before the course, taking statistics concurrently, or not satisfying the statistics prerequisite (Table 2). Seventy-eight additional students enrolled in introductory biology did not complete either the presurvey or postsurvey, and they were not included in the study sample. We compared student GPAs of the 186 survey completers and 78 noncompleters to determine whether the study sample was representative of the student population completing the course.

We retested the 2006 cohort 1 yr after completion of their postsurvey, in a sophomore-level required biology course (Biology 215, ecology and evolution). The survey was given in-class during the last week of the semester. This resulted in a sample of 30 students. The sample size was small because a large number of students take

**Table 1.** Design of statistics survey and student responses

Item	Category	Survey question	Correct responses presurvey (%) <sup>a</sup>	Correct responses postsurvey (%) <sup>a</sup>
1	Identification	Histogram	67 (36)	127 (69)
2	Identification	Scatterplot	183 (98)	184 (99)
3	Identification	Stem-and-leaf plot	149 (80)	184 (99)
4	Graphical tools	Box plot	113 (61)	142 (76)
5	Graphical tools	Histogram	99 (53)	116 (62)
6	Graphical tools	Scatterplot	140 (75)	157 (84)
7	Experimental analysis	Quartiles	162 (88)	175 (94)
8	Experimental analysis	$p$ value	66 (35)	128 (69)
9	Experimental analysis	Causation vs. correlation	96 (52)	126 (68)
10	Data evaluation	Causation/correlation	87 (47)	90 (48)
11	Data evaluation <sup>b</sup>	$p$ value	65 (35)	81 (43)

<sup>a</sup> Total students participating in survey  $n = 186$ .

<sup>b</sup> Includes both answer choices C and E.

**Table 2.** Number and gender distribution of study participants

	Male	Female	Total
Previous statistics	63	67	130
Concurrent statistics	17	14	31
No statistics	19	6	25
Total	99	87	186

this course out of sequence or pass out of it with advanced placement or transfer credit.

### Data Analysis

Pre- and postsurveys and demographic data for individual students were matched, and names and student IDs were removed from the data set before analysis. We calculated individual student normalized gains ( $g$ ), where  $g$  is the proportion of the actual score gains compared with the maximal gain possible, given the student's initial score;  $g = (\text{postscore} - \text{prescore}) / (100\% - \text{prescore})$ , for each student in the sample. Average normalized gain,  $G$ , defined as the mean of  $g$  values in a population, was then calculated for the following student groups: those who had completed a college-level introductory statistics course in a previous semester, those taking statistics concurrently, and students who had no college-level statistics preparation. We also calculated  $G$  for students who had completed a biology-themed section of introductory statistics, and we compared survey performance for male and female students. For students retested 1 yr after completing introductory biology,  $G$  was calculated using the student's initial presurvey score from the start of spring 2006 and the postsurvey score from the completion of spring 2007, representing a spacing of approximately 16 mo. The assumption of normality for the prescore, postscore, and normalized gain data sets was confirmed by skew and Kurtosis analysis (skew and Kurtosis  $< 1$ ). All statistical tests for data significance ( $t$  test, analysis of variance [ANOVA], analysis of covariance [ANCOVA]), including skew/Kurtosis calculations and corrections for simultaneous comparisons of groups (Tukey's or Bonferroni) were performed using Minitab version 15.1.1.0.

## RESULTS

We developed a survey to measure statistics learning gains in an introductory biology course that emphasized statistical analysis to determine whether statistics knowledge learned in introductory statistics could be either maintained or improved by extra practice and review in a nonstatistics course. The survey tested student recall and understanding of the concepts from introductory statistics that are necessary for basic data analysis in introductory biology. Of the 264 students completing introductory biology, 186 students (70%) who completed both the initial and final surveys were included in the study sample. The mean GPA was 0.35 points higher for study participants (3.36) than for those who did not complete the study (3.01). The difference was statistically significant ( $t$  test,  $p < 0.005$ ). The mean course grade for nonparticipants was also significantly lower (3.40 for participants, 2.71 for nonparticipants;  $t$  test,  $p < 0.005$ ). These data suggest that the study did not adequately represent students who are "strugglers" in the major. Their exclusion from the study seems to be due to lower attendance rates by struggling students; 98% of students who were excluded from the study sample failed to complete the post-

survey, which was given in class during the last week of classes for the semester.

The statistics survey consisted of questions in four categories: identification of graphs, use of graphical tools, analysis of experimental data, and evaluation (Table 1). The instrument was field tested with 186 biomedical science majors. We gauged question difficulty by the percentage of correct responses on the presurvey. Presurvey correct responses on individual items ranged from nearly 100% (item 2) to 35% (items 8 and 11). Kaplan and Saccuzzo (1997) suggest that instrument multiple-choice questions ideally have initial correct response averages about halfway between random guess and 100%. Because most of the survey items had either four or five answer choices, an initial correct response rate of ~60% would have been ideal. Thus, some questions that had an 80% or greater initial correct response rate had little utility in gauging learning (e.g., items 2, 3, and 7), although they were useful in demonstrating students' incoming knowledge. Other items (e.g., item 11) were likely too difficult to accurately measure student understanding of that particular concept.

The survey was reliable in measuring an increase in statistics knowledge in introductory biology. We saw a statistically significant improvement of scores on the survey of 24 percentage points between first and second administration of the instrument (from 57% correct to 71% correct;  $t$  test,  $p < 0.005$ ). Figure 1A indicates median pre- and postsurvey scores for students in all groups. All three groups showed statistically significant improvement in their average survey scores between the pre- and postsurvey (Figure 1B): 9 percentage points (from 64 to 73%) for students with previous statistics course work, 32 percentage points for those taking statistics concurrently (from 41 to 73%), and 19 points (from 43 to 62%) for those without statistics training. Each improvement was significant within individual groups ( $t$  test,  $p < 0.005$  in all instances). Confidence intervals indicated in Figure 1B include Bonferroni's correction for simultaneous confidence levels.

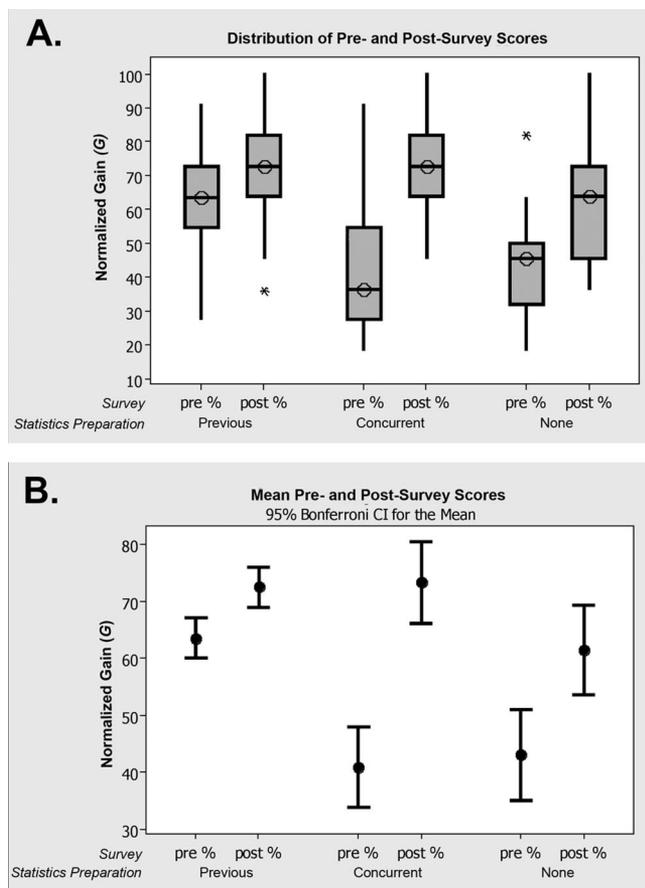
The normalized gain ( $G$ ) is an accepted measure used to quantify learning in science courses. It is commonly used in physics and astronomy learning studies (Hake, 1998; Hemenway *et al.*, 2002; Coletta and Phillips, 2005). The overall  $G$  for our study sample was 0.24 (SD = 0.44).  $G$  was calculated for groups with varied statistics preparation (Figure 2). Students who had completed introductory statistics in the prior semester showed modest gains ( $G = 0.16$ , SD = 0.46) on the survey. Both groups of students with no previous statistics at the beginning of the course showed stronger gains than the group with previous statistics. For students who took statistics concurrently,  $G = 0.53$  (SD = 0.26) while for students who took no statistics course at all,  $G = 0.3$  (SD 0.36). Although they posted lower scores on the presurvey, the mean postsurvey score for concurrent students was comparable with that of students who had completed the statistics prerequisite (Figure 1B). This suggests that students taking statistics concurrently were on par, by semester's end, with students who took statistics before entering biology. Concurrent students showed the greatest  $G$  of all groups (Figure 2A). It is not surprising that this group posted the largest gain of all three groups, because these students were receiving concurrent instruction in statistics in both biology and introductory statistics. The gains shown by these stu-

dents therefore include gains from the biology course and gains from the statistics course.

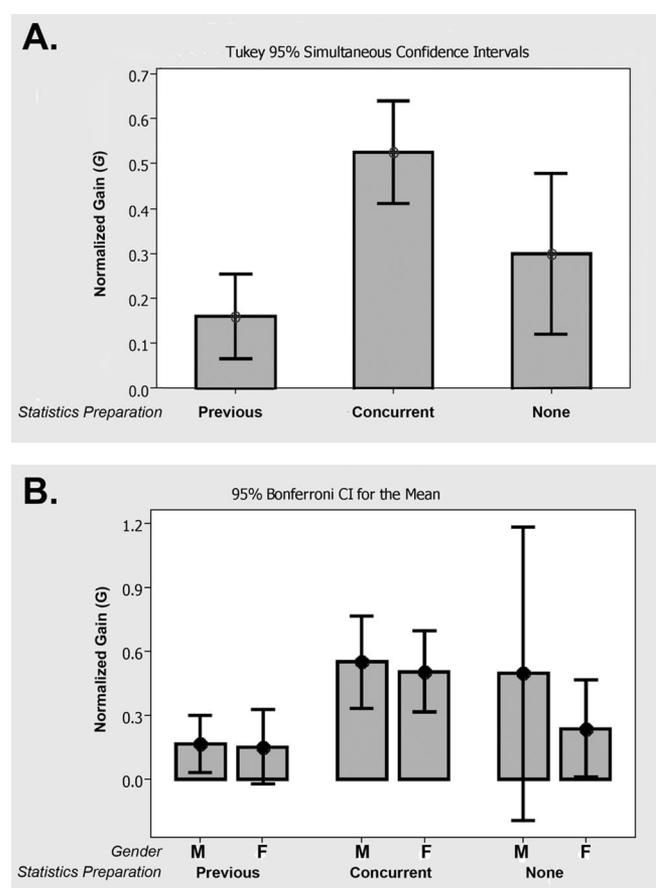
Of the students in the study sample, 53.2% were female, which was proportional to the percentage of women in the course over two semesters (264 students; 53.5% female). Similar numbers of males and females were in the different statistics preparation groups, with the exception of students without statistics course work (Table 2). We found no statistically significant difference in the performance of male and female students in this study (Figure 2B;  $G = 0.23$  [SD = 0.46] for female students and  $G = 0.25$  [SD = 0.41] for male students [ $t$  test,  $p = 0.719$ ]). Female students showed slightly higher average course grades (3.45 vs. 3.34) and GPA (3.41 vs. 3.31) than male students. These differences were not significant (ANOVA,  $p = 0.255$  for course grades,  $p = 0.222$  for GPA). The difference in learning gains between male and female students was still not significant even when these

slight course grade and GPA differences were accounted for via covariate (ANCOVA) analysis.

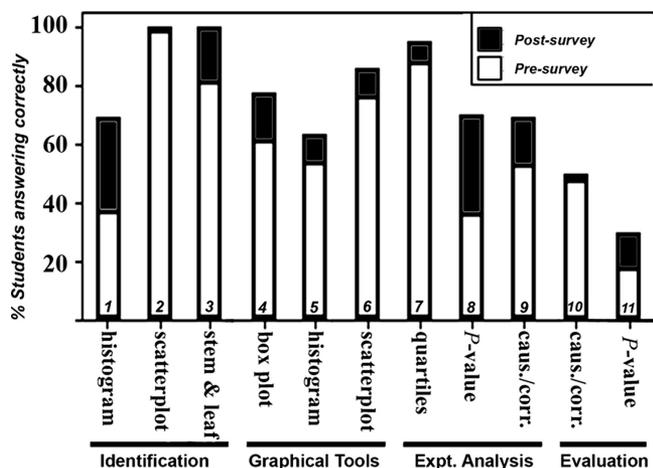
We designed individual survey items to track varying aspects of students' statistics knowledge. Survey items 1–3 tested whether students could recognize commonly used graphical tools (a histogram, scatterplot, and stem-and-leaf plot). Students had little difficulty identifying a scatterplot (item 2) even without having completed college statistics, posting 98% correct responses on the initial survey (Figure 3). Stem-and-leaf plots (item 3) were also easily identified, especially by semester's end (99% correct responses on final survey). Students struggled more with the distinction between a bar chart and a histogram (item 1), with 59% initially choosing "bar chart." The bar chart question is notable both for identifying a point of student confusion, and as a more general caveat for instrument design.



**Figure 1.** Pre- and postsurvey scores for groups with different levels of statistics course work preparation. Score increases were significant for students who completed statistics prior (Previous) to course enrollment ( $N = 130$ ;  $t$  test,  $p < 0.005$ ), students enrolled in statistics concurrently ( $N = 31$ ;  $t$  test,  $p < 0.005$ ), and students with no formal statistics (None) training ( $N = 25$ ;  $t$  test,  $p < 0.005$ ). (A) Boxplot showing score distribution of pre- and postscores for all groups. Outliers are indicated by asterisk (\*). Median scores are indicated by ○. (B) Mean pre- and postscores for all, with 95% Bonferroni's confidence intervals. Mean scores are indicated by ●. Skew < 0.90, Kurtosis < 0.70 for all populations.



**Figure 2.** (A) Average normalized gains ( $G$ ; ○) by students having completed introductory statistics (Previous,  $N = 130$ ), taking introductory statistics concurrently (Concurrent,  $N = 31$ ), and having no formal statistics instruction (None,  $N = 25$ ). Comparisons via ANOVA, with Tukey 95% simultaneous confidence intervals indicated. (B)  $G$  (●) by male and female students in each statistics preparation category. The 95% confidence intervals with Bonferroni's corrections are indicated. Previous statistics: male,  $N = 63$ ; female,  $N = 67$ . Concurrent statistics: male,  $N = 17$ ; female,  $N = 14$ . No statistics: male,  $N = 19$ ; female,  $N = 6$ . Skew < 0.90, Kurtosis < 0.70 for all populations.



**Figure 3.** Percentage of students answering correctly on statistics pre- and postsurvey items. Item number (1–11) is shown in italics at the base of each bar.  $N = 186$  for each item.

The confusion on this question is understandable, because histograms are technically a special case of a bar chart, and they look much alike. However, the distinction between the two is important, because both categorical data (properly displayed on bar charts) and continuous data (properly displayed on histograms) are commonly encountered in biology. Our students are taught that these are two separate types of graphs (the text used in our introductory statistics course (Moore and McCabe, 2003) explicitly distinguishes the two), but on the initial survey, only 36% chose histogram as the correct answer when bar chart was also an answer choice. However, after using histograms in the introductory biology course, the number of students choosing “histogram” as the correct answer on item 1 nearly doubled (from 67 to 127). Thus, although the bar graph and histogram answer choices are both technically correct and the question needs to be refined, the data have indicated to us that further instruction, giving biology students a more thorough understanding of histograms and bar charts, is beneficial.

Three items of the survey (items 4–6) asked students to choose a graph appropriate to a given data set (box plot, histogram, or scatterplot). The items had initial correct response rates between 53 and 75%, and they showed an increase in correct responses between 12 and 26% on the posttest (Figure 3). These questions indicated that students showed gains in their ability to choose the correct graphical tools to appropriately display data.

The last five survey items represented higher (analysis and evaluation) levels of Bloom’s taxonomy. Three questions required students to analyze experimental data (items 7–9). Most students (88% on the presurvey) correctly analyzed quartile data (item 7). Modest gains (31%) were seen in the ability of students to identify a problem in confusing causation with correlation (item 9). The largest score increase was observed on survey item 8, where students were asked to interpret the meaning of the  $p$  value statistics (66–128 correct, a 94% gain).

The lowest overall correct responses on the survey were observed with two evaluation questions, regarding  $p$  value and determining a causative relationship between variables (items 10 and 11). The inherent difficulty of this level of thinking was borne out by correspondingly low correct response rates. The question about causative relationships was answered correctly on both the pre- and postsurveys by ~50% of study participants (item 10). Student confusion regarding  $p$  value is not surprising. There are at least 13 different misconceptions of  $p$  value that have been documented in the literature across mathematics education, statistics, and psychology disciplines (Lane-Getaz, 2005). Research indicates that 61% of lower-division undergraduates believe a low  $p$  value indicates that the null hypothesis is true, and more than half believe a large  $p$  value means that data are statistically significant (Lane-Getaz, 2005).

Item 11 also elicited a low initial correct response rate. The question was problematic, because answer choice E was ambiguous and could be interpreted as a correct response in addition to answer choice C; if both answer choices are counted, 43% of students answered item 11 correctly on the posttest (35% on pretest). Given that the two choices represent 40% of the answer choices on this question, this suggests student answers were essentially random on both pre- and postsurveys. A suggested revision for the question is shown as “Revised 11” in the Supplemental Material.

Student responses to item 8 better indicate that supplemental instruction and practice in statistics is improving statistics knowledge of biology students. In this item, students were asked to judge whether there was a statistically significant difference between two data sets with a  $p$  value of 0.06, and they were also asked to justify their answer in a short statement. Students were scored as providing a correct answer if the reasoning they provided was based on sound statistical principles. Thus, a student could argue the two populations were different (e.g., if an  $\alpha$  of 0.10 was assumed), or they could argue the data were not significant because the  $p$  value was above an assumed  $\alpha$  of 0.05. On the presurvey, only 35% of students gave reasonable answers, whereas that percentage almost doubled to 69% by the post-survey. For the class as a whole, it seems that the basic understanding of how a  $p$  value is used improved, even if students could not precisely define the  $p$  value itself.

We compared learning gains between students who had previously taken a specialized statistics section emphasizing biological examples and students who completed the same introductory statistics course without biological emphasis. The  $G$  for students in the biology statistics section was more than double ( $G = 0.25$ ) that of nonbiology section statistics students that year ( $G = 0.10$ ), and also greater than  $G$  of students who completed regular introductory statistics courses in the following year (Table 3). This finding could not be shown to have statistical significance (ANOVA,  $p = 0.584$ ), but it may warrant further study. These preliminary data are particularly interesting because it suggests that teaching statistics in a manner that ties the material to a student’s subject of interest results in better long-term retention of statistical knowledge, and it may increase students’ ability to use statistical analysis within their own discipline.

**Table 3. Gains for students taking regular and biology-themed introductory statistics**

Statistics preparation	n	G <sup>a</sup>	G SD
Statistics (2005)	44	0.10	0.55
Statistics (2006)	53	0.16	0.54
Biology-themed statistics (2005)	16	0.25	0.40

<sup>a</sup> Differences were not statistically significant (ANOVA with Tukey's correction for multiple pairwise comparisons,  $p = 0.584$ ; population skew =  $-0.86$ , Kurtosis =  $0.44$ ).

To determine whether the difference in  $G$  was due to an inherent difference in average ability between the students in the two groups, we compared the mean GPA for the two groups. There was no statistically significant difference between the mean GPA of the biology statistics group and the regular statistics group ( $t$  test,  $p = 0.557$ ).

Students who were surveyed in spring 2006 were surveyed again 1 yr later, in a required sophomore-level biology course (spring 2007). Students who completed both pre- and postsurveys in 2006, and the survey again in 2007 ( $N = 30$ ), showed a slightly higher  $G$  after three semesters ( $G = 0.19$ ,  $SD = 0.40$ ), than after one semester ( $G = 0.16$ ,  $SD = 0.50$ ). These values were not statistically distinguishable ( $t$  test,  $p = 0.839$ ), suggesting retention or slight improvement in their level of statistics knowledge with progression through the curriculum.

## DISCUSSION

When our department faculty first began to reform the introductory biology curriculum using an inquiry-based learning model, we noted that students had difficulty with quantitative analysis and exhibited anxiety regarding statistical data analysis. Statistics anxiety, well documented in the classroom (e.g., Onwuegbuzie *et al.*, 1997), seems to stem from a more generalized math phobia, fear of failure, and the student's perception that statistics is disconnected from his or her real-world experience (Pan and Tang, 2005). We recognized that to integrate statistics into the biology curriculum, we would need to help our students overcome their own statistics anxiety.

Starting in 2006, we modified our introductory biology course to provide students with supplementary instruction in statistics, followed by the opportunity for students to practice statistical analysis of their own data generated in inquiry-based labs (Figure 4). Work in both statistics (e.g., Magel, 1996; Moore, 1997; Delucchi, 2006) and physics education research (e.g., Hake, 1998) suggests that students' analytical skills in a course improve most when students have the opportunity to practice problem solving during the semester. When statistics is emphasized in a biological or medical course context (e.g., Bahn, 1970; Peterson, 2000), student skills also seem to improve, although little work has been done on quantifying statistics learning within a biology curriculum. We therefore designed an 11-item statistics survey to quantify the effects of the statistics interventions in the biology course.



**Figure 4.** Students work in teams to design experiments and to collect and analyze data.

The normalized gain measure ( $G$ ) has been used in high school and college physics courses to measure the effectiveness of a course in promoting conceptual understanding. Hake (1998) reported a  $G$  of 0.23 for traditional physics (mechanics) courses using a standardized test of mechanics reasoning, the Force Concept Inventory (FCI). Higher  $G$  values (0.48–0.60) on the FCI for courses that emphasized problem-solving skills also have been reported previously (Coletta and Phillips, 2005). Because we were examining the learning of statistics within a biology course, we had expectations of fairly modest gains in increasing the conceptual understanding of statistics. Although a standard, published inventory was not used in this study, a  $G$  value of 0.25, similar to  $G$  values obtained using published instruments, suggests that the survey instrument was appropriate to the material being tested.

The survey was designed so that students who had completed elementary statistics should have been able to answer a high percentage of the questions correctly without studying. Overall, students scored well enough on the pretest (58% average) to allow improvement, but not so low as to indicate that the survey was unreasonably difficult. All student groups (those who had completed statistics, those taking statistics concurrently, and students with no formal statistics course work) showed gains in statistics knowledge by the end of the semester.

It has been reported that girls and young women trail behind their male peers in mathematics achievement test performance starting at about high school (Kiefer and Sekaquptewa, 2007). Other evidence suggests that female students learn less than their male peers in college science courses and underperform on science tests (e.g., Hake, 1998; McCullough, 2004). We were therefore interested in deter-

mining whether there were gender differences in initial performance and in learning gains between male and female students on the statistics survey.

No statistically significant differences were found in the performance of male and female students on the survey (Figure 2). Given that female performance is often found to be lagging behind that of male students in science disciplines, it is important that no statistically significant differences in learning, as determined by this survey instrument, were detected. It may be that learning differences disappear as women gain parity in numbers in the classroom. The rate of female participation in both the study and the introductory biology courses (~53%) was similar to rates seen nationally. In 2001, 57% of biological science bachelor degree recipients were women, although the percentage of women in biology drops off in graduate school and into higher levels of academia (Sible *et al.*, 2006).

We did observe that women consistently showed slightly lower *G* than men in all categories, even when both their average course grade and GPA were higher. Recent work indicates that although girls outperform boys in classroom mathematics work, they have until very recently trailed behind boys on standardized achievement tests, perhaps due to “stereotype threat”—the conscious or subconscious influence of negative stereotypes in our culture regarding girls’ mathematics competence (Steele *et al.*, 2002; Kenney-Benson *et al.*, 2006). It may be that stereotype threat somewhat affected women’s own assessment of competence during what was perceived as a “surprise math quiz” and caused them, as a group, to slightly underperform compared with the male students.

We made a preliminary observation that a statistics course emphasizing biological examples seemed to contribute to gains in understanding of statistics in a subsequent biology course (Table 3). With funding from a Howard Hughes Medical Institute education grant, we were able to offer, in fall 2005, one section of introductory statistics that heavily emphasized biological examples. The *G* of the student group that had previously been exposed to statistics in a biological context via this section was 2.5 times greater than the *G* of students who took an identical course in a section that used few, if any, biology examples.

Interestingly, survey prescores for both groups were similar. It may be that although all students initially have difficulty when presented with statistics in a biology course, it is the students who have had previous exposure to the use of statistics in biology who retain more of this knowledge upon second exposure. Berger *et al.* (1999) suggest that material that has been learned, and subsequently forgotten, can be reactivated with a minimal corrective intervention. Material so retrieved can then remain accessible to the student for years in the future. In our case, it seems that using statistics in introductory biology may act as a corrective intervention, where the memories that have been lost (in this case, of the use of statistics in biology) are quickly recovered when a student re-encounters the material in the biology classroom. Because gains on the survey were stronger for students completing statistics with a biology emphasis than for students completing general statistics, it may be that teaching statistics in a more integrated manner in the first place is helpful.

It must be noted that our sample size for this analysis was very small, and this phenomenon requires further investigation. Unfortunately, we were unable to institutionalize a biology-themed statistics course at MSU, and we have not been able to confirm this trend. Barriers to institutionalization of the course were twofold. One issue is the department-based organizational structure of the university. No agreement could be reached across the mathematics and biology departments as to which department would pay for and staff the course in future years. The second barrier was more philosophical. Statistics is a service course taught by the mathematical sciences department for majors across campus, including engineering, physics, business, chemistry, and many biological sciences. It was felt that if efforts were made to make special sections for biology, it would lead to a demand for special sections for all majors and would make scheduling and administration of this already complex course (approximately 40 sections and 1600 students per year) unmanageable.

For statistical learning to continue to be useful to students after the completion of introductory statistics, this knowledge needs to be stabilized to allow long-term access. Re-testing of our students 1 yr after the completion of introductory biology indicated that they had retained their statistics knowledge at levels at or above where they were after the completion of introductory biology. Research on the nature of memory indicates that repeated exposure to material is useful in long-term retention of knowledge (Bjork, 1988; Bahrick and Hall, 1991a,b). Bahrick and Hall (2005) show that practice sessions spaced apart are critical for long-term retention of knowledge. This may be because intervals between use of the knowledge means that retrieval failures will occur, which help students identify gaps in knowledge or understanding that can then be rectified. It is possible that students initially experienced retrieval failures upon entering biology (students posted low survey scores at the start of the biology course), followed by relearning during the semester. Such relearning may have allowed students to retain their statistics knowledge over longer time frames.

Beyond looking at overall trends on the survey, analysis of student responses to individual questions has been enlightening. For example, survey results indicated that the concept of *p* value is difficult for our students, but also that students improved their understanding of *p* value in the biology course. Alternatively, there was almost no improvement on item 10 (an evaluative question on the difference between causation and correlation), indicating that we did not teach this concept well and need to make curricular changes. The essentially random responses on item 11 as administered indicate that the question was poorly crafted and not valuable in assessing student learning. This item needs to be improved on future generations of the instrument, ideally incorporating the most common misconceptions as foils. The results from this initial pilot of the survey will also help to improve the instrument, and a better understanding of the strengths and weaknesses in student understanding of statistics will allow us to better tailor course content to the needs of our students in future semesters.

A final issue of note is the problem in classroom research of capturing strugglers—students who are not fully engaged in the course and less likely to complete the study than their more successful peers. In this study, pre- and postsurvey

data were obtained for only 70% of students completing the course, and nonparticipating students had significantly lower GPAs and course grades than participating students. Obtaining data on nonparticipating students is a problem inherent to classroom research; the population is nonrandom and classroom circumstances must be considered in the research design. Possible solutions include attempting to catch students in subsequent class periods or providing the opportunity for students to take posttests at their convenience (e.g., online). However, both of these methods will likely suffer from the same lack of student participation, and additionally add variables (e.g., time on task or advance notice) that may confound the data. Embedding the post-survey in a final exam is most likely to capture all the students who complete the course. However, even this tactic may not be appropriate, depending on the scope of the instrument or the design of the course, and also it will not capture students who fail to finish the class. Accounting for the influence of struggling students on the course population should be given the fullest consideration in study design; data from such studies will be richer and better indicate how courses can be improved to serve every student who walks into the classroom.

## CONCLUSIONS

Students in an introductory biology course showed statistically significant gains in their understanding of statistics when given the means and opportunity to practice data gathering and analysis of biological experiments. Learning gains were measured using a new instrument designed specifically to measure the students' knowledge of statistics in the biology classroom. The work provides a model for strengthening the statistical analysis skills of biology majors.

## ACKNOWLEDGMENTS

We are grateful to the students of Biology 213 and 215 for support of the study and positive suggestions for improving the courses. I thank Clarissa Dirks (Evergreen College) and Matthew Cunningham (University of Washington) for project brainstorming and initial review of the statistics instrument; Alex Dimitrov for initial outline and final review of the Statistics Primer used in the introductory biology sequence; Gwen Jacobs, Cathy Cripps, Steven Kalinowski, and Mark Taper for Biology 213 and 215 course design and help in implementing the study; and Martha Peters and Cali Morrison for aid in completing the study. This research was supported in part by a grant to MSU from the Howard Hughes Medical Institute through the Undergraduate Science Education Program, and it was performed with the review and approval of the Institutional Review Board for the Protection of Human Subjects at MSU Bozeman (exempt status).

## REFERENCES

A'Brook, R., and Weyers, J.D.B. (1996). Teaching of statistics to UK undergraduate biology students in 1995. *J. Biol. Educ.* 30, 281–288.

Bahn, A. K. (1970). Goals and methods of evaluating the teaching of biostatistics to medical students. *Arch. Environ. Health* 21, 559–568.

Bahrnick, H. P., and Hall, L. K. (1991a). Preventive and corrective maintenance of access to knowledge. *Appl. Cogn. Psychol.* 5, 1–18.

Bahrnick, H. P., and Hall, L. K. (1991b). Lifetime maintenance of high school mathematics content. *J. Exp. Psychol. Gen.* 120, 20–33.

Bahrnick, H. P., and Hall, L. K. (2005). The importance of retrieval failures to long-term retention: a metacognitive explanation of the spacing effect. *J. Mem. Lang.* 52, 566–577.

Berger, S. A., Hall, L. K., and Bahrnick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *J. Exp. Psychol. Appl.* 5, 438–447.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In: *Practical Aspects of Memory: Current Research and Issues*, Volume 1, *Memory of Everyday Life*, ed. M. M. Gruneberg, P. E. Morris, and R. N. Sykes, New York: Wiley, 396–401.

Bloom, B. S. (ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*, New York: Longman.

Budnitz, N., and Guentensberger, T. (2001). Find Your Peanut. Center for Inquiry-based Learning. Duke University, Durham, NC. [www.biology.duke.edu/cibl/exercised/find\\_your\\_peanut.htm](http://www.biology.duke.edu/cibl/exercised/find_your_peanut.htm) (accessed 10 July 2007).

Coletta, V. P., and Phillips, J. A. (2005). Interpreting FCI scores: normalized gain, preinstruction scores and scientific reasoning ability. *Am. J. Phys.* 73, 1172–1182.

Delucchi, M. (2006). The efficacy of collaborative learning groups in an undergraduate statistics course. *Coll. Teach.* 54, 244–248.

Garfield, J. B. (2003). Assessing statistical reasoning. *Stat. Educ. Res. J.* 2, 22–38.

Garfield, J., Hogg, B., Schau, C., and Whittinghill, D. (2002). First courses in statistical science: the status of educational reform efforts. *J. Stat. Educ.* 10(2). [www.amstat.org/publications/jse/v10n2/garfield.html](http://www.amstat.org/publications/jse/v10n2/garfield.html) (accessed 12 November 2007)

Guidelines for the Assessment and Instruction in Statistics Education (2005). American Statistical Association Guidelines for the Assessment and Instruction in Statistics Education (GAISE) Project, February, 2005. [www.amstat.org/education/gaise/](http://www.amstat.org/education/gaise/) (accessed 12 November 2007).

Hake, R. R. (1998). Interactive-engagement versus traditional methods; a six thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* 66, 67–74.

Hemenway, M. K., Straits, W. J., Wilke, R. R., and Hufnagel, B. (2002). Educational research in an introductory astronomy course. *Innov. High. Educ.* 26, 271–280.

Horgan, G. W., Elston, D. A., Franklin, M. F., Glasbey, C. A., Hunter, E. A., Talbot, M., Kempton, R. A., McNicol, J. W., and Wright, F. (1999). Teaching statistics to biological research scientists. *Statistician* 48, 393–400.

Kaplan, R. M., and Saccuzzo, D. P. (1997). *Psychological Testing: Principles, Application and Issues*, 6th ed., Belmont, CA: Wadsworth Publishing.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., and Patrick, H. (2006). Sex differences in math performance: the role of children's approach to schoolwork. *Dev. Psychol.* 42, 11–26.

Kiefer, A. K., and Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification and math-related outcomes: a prospective study of female college students. *Psychol. Sci.* 18, 13–18.

Lane-Getaz, S. J. (2005). Summary of P-value survey research. 2005 United States Conference on Teaching Statistics (USCOTS), 19–21 May 2005, The Ohio State University, Columbus, OH. Columbus, OH: CAUSE, Department of Statistics, The Ohio State University. [www.causeweb.org/uscots/uscots05/spotlight/files/R6.pdf](http://www.causeweb.org/uscots/uscots05/spotlight/files/R6.pdf).

Luckie, D. B., Maleszewski, J. J., Loznak, S. D., and Krha, M. (2004). Infusion of collaborative inquiry throughout a biology curriculum

- increases student learning: a four-year study of “teams and streams.” *Adv. Physiol. Educ.* 287, 199–209.
- Magel, R. C. (1996). Increasing participation in large introductory statistics classes. *Am. Stat.* 50, 51–56.
- Maret, T. J., and Ziemba, R. E. (1997). Statistics and hypothesis testing in biology: teaching students the relationship between statistical tests and scientific hypotheses. *J. Coll. Sci. Teach.* 26, 283–285.
- McCullough, L. (2004). Gender, context and physics assessment. *J. Int. Womens Stud.* 5, 20–30.
- Moore, D. S. (1997). New pedagogy and new content: the case of statistics. *Int. Stat. Rev.* 65, 123–165.
- Moore, D. S., and McCabe, G. P. (2003). *Introduction to the Practice of Statistics*, 4th ed., New York: Freeman and Co.
- National Research Council (2003). *BIO 2010, Transforming Undergraduate Education for Future Research Biologists*. Committee on Undergraduate Education to Prepare Research Scientists for the 21st Century, Washington, DC: National Academies Press.
- Onwuegbuzie, A. J., DaRos, J., and Ryan, J. (1997). The components of statistics anxiety: a phenomenological study. *Focus Learn. Probl. Math.* 19, 11–35.
- Pan, W., and Tang, M. (2005). Students’ perceptions on factors of statistics anxiety and instructional strategies. *J. Instr. Psychol.* 32, 205–214.
- Peterson, C. E. (2000). An experimental project approach to biology. *J. Coll. Sci. Teach.* 30, 162–165.
- Sible, J. C., Wilhelm, D. E., and Lederman, M. (2006). Teaching cellular and molecular biology for gender equity. *CBE Life Sci. Educ.* 5, 227–238.
- Steele, C. M., Spencer, S. J., and Aronson, J. (2002). Contending with group image: the psychology of stereotype and social identity threat. In: *Advances in Experimental Social Psychology*, Vol. 34, ed. M. Zanna, New York: Academic Press, 379–340.